Prostate cancer risk prediction using the novel versions of the ERSPC and PCPT risk calculators: Independent validation and comparison in a contemporary European cohort

Poyet, Cédric ; Nieboer, Daan ; Bhindi, Bimal ; Kulkarni, Girish S ; Wiederkehr, Caroline ; Wettstein, Marian S ; Largo, Remo ; Wild, Peter ; Sulser, Tullio ; Hermanns, Thomas

Abstract: OBJECTIVES: To externally validate and compare the two novel versions of the ERSPC-prostate cancer (PCa) risk-calculator (RC) and PCPT-RC. PATIENTS AND METHODS: All men who underwent a transrectal prostate biopsy in a European tertiary care centre between 2004 and 2012 were retrospectively identified. The probability of detecting PCa and significant PCa (Gleason score 7) was calculated for each man using the novel versions of the ERSPC-RC (DRE-based version 3 / 4) and the PCPT-RC (version 2.0) and compared with the biopsy results. Calibration and discrimination were assessed using the calibration slope method and the area under the receiver operating characteristic curve (AUC), respectively. Additionally, decision curve analyses were performed. RESULTS: Of 1996 men, 483 (24%) were diagnosed with PCa and 226 (11%) with significant PCa. Calibration of the two RCs was comparable, although the PCPT-RC was slightly superior in the higher risk prediction range for any and significant PCa. Discrimination of the ERSPC- and PCPT-RC was comparable for any PCa (AUCs: 0.65 vs. 0.66), while the ERSPC-RC was somewhat better for significant PCa (AUCs: 0.73 vs. 0.70). Decision curve analyses revealed a comparable net benefit for any PCa and a slightly greater net benefit for significant PCa using the ERSPC-RC. CONCLUSIONS: In our independent external validation, both updated RCs showed less optimistic performance compared to their original reports particularly for the prediction of any PCa. Risk prediction of significant PCa, which is important to avoid unnecessary biopsies and reduce overdiagnosis and overtreatment, was better for both RCs and slightly superior using the ERSPC-RC.

DOI: https://doi.org/10.1111/bju.13314

Prostate cancer risk prediction using the novel versions of the ERSPC and PCPT risk calculators: Independent validation and comparison in a contemporary European cohort

Cédric Poyet[a*], Daan Nieboer[b], Bimal Bhindi[c], Girish S. Kulkarni[c], Caroline Wiederkehr[a], Marian S. Wettstein[a], Remo Largo[a], Peter Wild[d], Tullio Sulser[a] and Thomas Hermanns[a]

[a]*Department of Urology, University Hospital Zürich, University of Zürich, Zürich, Switzerland*
[b]*Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands*
[c] *Department of Surgery, Division of Urology, University Health Network, University of Toronto, Toronto, Canada*
[d] *Institute of Surgical Pathology, University Hospital Zurich, Zurich, Switzerland*

*Corresponding author:
Department of Urology
University Hospital Zürich
Frauenklinikstrasse 10
8091 Zürich, Switzerland
Phone: +41-44-255-1111
Fax: +41-44-255-4555
email: cedric.poyet@usz.ch

Email addresses:
Cédric Poyet: cedric.poyet@usz.ch
Daan Nieboer: d.nieboer@erasmusmc.nl
Bimal Bhindi: bimal.bhindi@mail.utoronto.ca
Girish S. Kulkarni: girish.kulkarni@uhn.ca
Caroline Wiederkehr: caroline.wiederkehr@gmx.ch
Marian S. Wettstein: marian.wettstein@me.com
Remo Largo: remo.largo@usz.ch
Peter Wild: peter.wild@usz.ch
Tullio Sulser: tullio.sulser@usz.ch
Thomas Hermanns: thomas.hermanns@usz.ch

ABSTRACT

Objectives: To externally validate and compare the two novel versions of the ERSPC- prostate cancer (PCa) risk-calculator (RC) and PCPT-RC.

Patients and methods: All men who underwent a transrectal prostate biopsy in a European tertiary care centre between 2004 and 2012 were retrospectively identified. The probability of detecting PCa and significant PCa (Gleason score ≥7) was calculated for each man using the novel versions of the ERSPC-RC (DRE-based version 3 / 4) and the PCPT-RC (version 2.0) and compared with the biopsy results. Calibration and discrimination were assessed using the calibration slope method and the area under the receiver operating characteristic curve (AUC), respectively. Additionally, decision curve analyses were performed.

Results: Of 1996 men, 483 (24%) were diagnosed with PCa and 226 (11%) with significant PCa. Calibration of the two RCs was comparable, although the PCPT-RC was slightly superior in the higher risk prediction range for any and significant PCa. Discrimination of the ERSPC- and PCPT-RC was comparable for any PCa (AUCs: 0.65 vs. 0.66), while the ERSPC-RC was somewhat better for significant PCa (AUCs: 0.73 vs. 0.70). Decision curve analyses revealed a comparable net benefit for any PCa and a slightly greater net benefit for significant PCa using the ERSPC-RC.

Conclusions: In our independent external validation, both updated RCs showed less optimistic performance compared to their original reports particularly for the prediction of any PCa. Risk prediction of significant PCa, which is important to avoid unnecessary biopsies and reduce overdiagnosis and overtreatment, was better for both RCs and slightly superior using the ERSPC-RC.

Keywords

Prostate cancer, biopsy, prostate-specific antigen, nomograms, decision aids

Introduction

The widespread use of prostate-specific antigen (PSA) testing led to an increased prostate cancer (PCa) detection rate and a shift from advanced to earlier disease states at diagnosis [1]. Earlier PCa treatment is thought to have contributed to the reduced incidence of advanced and metastatic disease and PCa-specific mortality [2, 3].

Classically, a serum PSA value above a pre-defined threshold or a suspicious digital rectal examination (DRE) triggered the decision to perform a prostate biopsy. However, the limited sensitivity and specificity of PSA and DRE alone resulted in a high number of negative, thus unnecessary biopsies [4]. Furthermore, PSA-based screening also increased the detection of potentially indolent PCa, which often does not become clinically relevant even when left untreated [5, 6]. Diagnosis and subsequent treatment of insignificant PCa can be regarded as overdiagnosis and overtreatment [5]. In screen-detected cohorts, the estimates of PCa overdiagnosis range from 23% to 44% [1].

Numerous nomogram-based PCa risk-calculators (RC) have been developed to improve risk-prediction and to overcome the problems of unnecessary biopsies, overdiagnosis and overtreatment [7-10]. It has been shown that these RCs are more accurate in predicting the likelihood of PCa detection than PSA and DRE alone [7, 11].

The ERSPC-RC and the PCPT-RC are two well-known PCa RCs based on data from the Dutch arm of the European Randomized Study for Screening of Prostate Cancer (ERSPC, 15’758 men) and the placebo arm of the Prostate Cancer Prevention Trial (PCPT, 5’519 men), respectively [8, 9]. Both RCs have been extensively validated in independend cohorts with distinct variation in performance among the different validation cohorts [12-16]. Modifications of the original versions of both RCs have

constantly been performed to improve their performance. Recently, novel versions of both RCs have been launched and promising results in development and validation cohorts have been reported in their original publications [17, 18]. So far, only one small independent validation study of these novel versions has been performed. However, to date, an independent validation and comparison in an adequately large patient cohort has not been conducted [19]. The aim of the present investigation was to independently validate and compare the performance of the two updated RCs.

Patients and methods

All men who underwent a transrectal ultrasound (TRUS)-guided prostate biopsy between January 2004 and July 2012 in a European tertiary care academic centre, were retrospectively identified. Pre-biopsy clinical and pathological data were obtained from electronic medical charts. Patients were excluded if they were older than 75 years, if their PSA was >50µg/l or if they had a previous positive biopsy (i.e. under active surveillance). Men with previous negative biopsies were included. This study was approved by the local ethics committee (approval no.: StV-Nr.0133/2012).

Generally, a prostate biopsy was considered if serum PSA was 2.5µg/l or greater or if a DRE was abnormal. Before 2007 either six- or eight-core biopsies and after 2007 only twelve-core biopsies were performed. Histological examination took place in our Institute of Surgical Pathology.

The probabilities of detection of PCa and significant PCa (Gleason Score ≥7) were calculated for each patient using the novel DRE-based ERSPC-RC (www.prostatecancer-riskcalculator.com) [8, 18] and the novel version of the PCPT-RC (version 2.0; http://deb.uthscsa.edu/URORiskCalc/Pages/calcs.jsp) [17]. The calculated probabilities were compared with the actual biopsy results for the entire cohort.

The novel ERSPC-RC is based on three variables (PSA, prostate volume and suspicious DRE; Table 1). The variable prostate volume is a DRE-based estimate trichotomized into three volumes (25ml, 40ml and 60ml). For our analysis, TRUS-measured prostate volume was used and trichotomized according to the predifined volumes of the RC (TRUS volume <30ml = 25ml, 30-50ml = 40ml and ≥50ml = 60ml). Risk calculation for biopsy-naïve

patients was done using the ERSPC-RC version 3+DRE. For patients with previous negative biopsies the RC version 4+DRE was used.

The novel PCPT-RC 2.0 is based on seven variables (Table 1) and gives separate predictions for not being diagnosed with PCa (no PCa) and for being diagnosed with insignificant PCa (Gleason score 6) and significant PCa. To compare the PCPT-RC 2.0 with the ERSPC-RC the prediction results had to be dichotomized: the risk of being diagnosed with any PCa was calculated from significant or insignificant PCa versus no PCa and the risk of being diagnosed with significant PCa was calculated from significant versus no or insignificant PCa.

Statistical analyses were performed using R statistical software version 3.1.0 (http://www.r-project.org). Associations between clinical and pathological variables were assessed using the chi-square test for categorical variables and the Wilcoxon rank-sum test for continuous variables. All *p*-values <0.05 were considered statistically significant. Missing values (n=53 for prostate volume) were single-imputed using the mice algorithm [20].

Calibration and discrimination were assessed for each RC. Calibration refers to the agreement between the predicted and observed proportion of events. Calibration was assessed graphically using a calibration plot and calibration-in-the-large. In a calibration plot, the predicted probabilities are plotted against the observed probabilities, allowing to assess the extent of risk over- or underestimation. The predicted probability refers to the probability of PCa for a given patient calculated for each RC. Patients were divided into quintiles (each group n=399) of predicted risk. Calibration-in-the-large was calculated by fitting a logistic regression model with the linear predictor of the models as an offset. The resulting intercept indicates whether predictions are systematically too low or too high. The calibration slope is a measure of the average strength of the predictors in a prediction

model and was obtained by fitting a logistic regression model with the linear predictor of the model as only covariate.

Discrimination refers to the ability of a prediction model to distinguish between patients with and without an event (any PCa or significant PCa) and was quantified using the area under the receiver operating characteristic curve (AUC). The AUCs of the ERSPC-RC and PCPT-RC 2.0 were compared using the method of DeLong and colleagues [21].

Additionally, decision curve analyses (DCAs) were performed for the assessment of the net benefit according to different threshold probabilities at which one would consider to perform a biopsy [22]. A probability threshold in a decision curve is different to the predicted probability in a calibration curve and refers to the a-priori cut point at which the clinician would do a biopsy.

In order to assess whether the number of biopsy cores taken impacted the performance of the novel RCs we performed a sensitivity analysis for patients who had either 6-8-core biospies or 12-core biopsies.

Additionally we aimed to analyze whether the performance of the novel versions of the two RCs are superior to their previous versions. For the ERSPC-RC we were not able to perform this analysis because the previous version include the variable suspicious TRUS lesion, which was not available in our dataset. However, for the two older versions of the PCPT-RC (original RC-1.0, RC-1.0 plus prostate volume; for details see Supplemental Table 1) we were able to perform the above mentioned analyses to asses differences in risk prediction of the different versions.

Results

Of 2304 identified men, 308 were excluded due to age >75 years, PSA >50µg/l or previous positive biopsies, resulting in 1996 men eligible for the final analysis. Table 2 summarizes their baseline characteristics. Overall, 1151 men (58%) were biopsy-naïve. PCa was detected in 483 men (24%) of which 226 (47%) had significant disease. Men with PCa were significantly older, had smaller prostates and higher PSA values. Furthermore, they were more likely to have a positive family history, a suspicious DRE but less likely to have a previous negative biopsy (Table 2).

The predicted proportions of any PCa and significant PCa were comparable between the ERSPC-RC and PCPT-RC 2.0 and were close to the observed proportion (Table 3).

The calibration plots of the ERSPC-RC showed good calibration in the risk range between 0 and 0.4 for both outcomes (Figures 1A, 2A). The calibration plots for the PCPT-RC 2.0 showed good calibration for both outcomes over the whole prediction range (Figures 1B, 2B). The calibration slope of the ERSPC-RC was lower than one, indicating that the effects of the predictors were on average too strong (Table 4). We were not able to calculate the calibration slope for the PCPT-RC 2.0, as the RC is based on a multinomial logistic regression model instead of a binary logistic regression model.

The discriminative ability for the detection of any PCa was not significantly different between the ERSPC RC and the PCPT-RC 2.0 (0.65 vs. 0.66, DeLong test p = 0.39; Table 3). For PSA alone the AUC for the prediction of any PCa was 0.58. For significant PCa the AUC of the ERSPC RC was higher compared to the PCPT-RC 2.0 (0.73 vs. 0.70; Table 4). The difference between the two AUCs was statistically significant (DeLong test p=0.043). The AUC for the prediction of significant PCa for PSA alone was 0.65.

To validate the original three predictions of the PCPT-RC 2.0, the discriminative abilities for no PCa, significant PCa and insignificant PCa were also assessed. The analysis revealed AUCs of 0.71 (significant versus no PCa), 0.64 (significant versus insignificant PCa) and 0.59 (insignificant versus no PCa).

The DCAs revealed that both RCs provided a clinical net benefit in the threshold probability range between 18% to 40% for any PCa and in the range between 8% and 40% for significant PCa (Figures 3 and 4, respectively). The net benefit was comparable between the two RCs for any PCa and somewhat greater for significant PCa if the ERSPC-RC was used.

Our sensitivity analysis revealed a higher detection rate for any PCa (29.4% vs. 18.1%) and significant PCa (15.9% vs. 5.9%) in patients receiving a 12-core biopsy (Supplemental Table 2). Both RCs overestimated the risk for any and significant PCa in the 6-8 core group and underestimated it in the 12-core group. The AUCs for the ERSPC-RC were only slightly lower in the 12-core group compared to the 6-8 core group. For the PCPT-RC 2.0 the AUC for signifcant PCa was markedly lower in the 12-core group compared to the 6-8 core group (0.66 vs. 0.78).

The analyses of the previous versions of the PCPT-RC revealed that both previous RCs overestimated the proportion of men diagnosed with any or significant PCa (Supplemental Table 3). Calibration plots for both previous PCPT RC versions confirmed this observation showing constant overestimation for both outcomes (Supplemental Figure 1 and 2). The PCPT-RC 1.0 with volume showed a better discriminative ability for significant PCa (AUC: 0.74) than the novel PCPT 2.0 (Supplemental Table 3).

Discussion

Prostate cancer RCs are important prediction tools to overcome the problems that arise from the widespread use of PSA screening. The optimal RC would help to minimize the number of unnecessary biospies and thereby reduces complications (i.e. systemic infections, bleeding) and emotional stress associated with prostate biopsies [23, 24]. Furthermore it would reliably predict the risk of being diagnosed with PCa and would give accurate predictions of having insignificant or significant disease. Optimization of these tools and their implementation into patient counseling and clinical decision making can help to decrease overdiagnosis and eventually overtreatment of PCa. Two well known RC for PCa risk prediction are the ERSPC and the PCPT RCs.

The recent updates of the ERSPC and PCPT RCs have been performed to further improve their performance [17, 18]. The ERSPC-RC does not include the parameter suspicious TRUS lesion anymore but integrates prostate volume estimated by DRE instead of TRUS volume [18]. These changes make the RC much more convenient to use in clinical pratice. For the novel PCPT-RC 2.0 no variables were added or removed compared to the previous version. However, data from more than 1000 biopsies from the PCPT placebo arm were added to the original dataset and three instead of two biopsy outcomes are now calculated [17]. Despite these updates and modifications the overall performance of the two RCs in our dataset was not optimal.

The present study is the first to externally validate and compare the novel versions of the ERSPC-RC and PCPT-RC in a large patient cohort of almost 2000 patients. Both RCs showed better performance than PSA alone but less optimistic performance compared to their original reports (ERSPC-RC: AUC 0.65 vs. 0.77 for any PCA and 0.73 vs. 0.85 for significant PCa; PCPT-RC 2.0: AUC 0.64 vs. median 0.68 for significant vs. insignificant

PCA and 0.71 vs. median 0.74 for significant vs. no PCa) [17, 18]. Foley and colleagues have recently reported a similar performance of both RCs for the detection of any PCa [19]. However, their validation study was conducted in a relatively small cohort of only 337 patients and RC performance for significant PCa was not assessed. In previous studies evaluating and comparing the older versions of the ERSPC and PCPT RCs, the performance of the RCs was also less optimistic in the validation cohorts compared to the original reports [12-15].

It has previously been shown that AUCs of nomograms are often lower in independent validations [11]. Differences in the tested populations are likely to account for the differences in nomogram performance. Predictor effects have been shown to be different in a population setting compared to a clinical setting [25]. Men in our cohort were biospied after individualized screening either because of an elevated PSA or a suspicious DRE. In contrast, both RCs are based on calculations from data of randomized controlled trials with protocol-mandated biopsies. The ERSPC-RC is based on results of a study evaluating population-based mass screening. The use of different PSA thresholds (3.0µg/L versus 2.5µg/L) might be an additional explanation for the differences in performance of the ERSPC-RC. Our sensitivity analysis revealed that the higher number of biopsy cores taken in more than 50% of our patients had only a minor impact on the RC performance in our cohort. The PCPT-RC has been developed from biopsy results of the placebo arm of the PCPT. All men were biopsied independently of their PSA value. Differences in the ethnic composition of the two cohorts may also explain the different performance. Although only 3.3% of men in the PCPT were African Americans they accounted for a 2.6-fold higher risk of having significant PCa cmpared to Caucasians which constituted the entire population of our cohort [9]. In contrast to the ERSPC-RC, the performance of the PCPT-RC 2.0 in our cohort was negatively influenced by the higher number of biopsy cores taken, particularly

for significant PCa. In addition, our cohort had a relevant amount of pre-biopsied men. Although both RCs account for prior negative biopsies this aspect explains the rather low detection rate in our cohort compared to others [26, 27].

Although a direct comparison between the updated ERSPC-RC and PCPT-RC 2.0 is not possible due to the different underlying statistical models, the results of our analysis indicate a slightly superior performance of the ERSPC, particularly for the prediction of significant PCa. A recent metaanalysis evaluating the older versions of the two RCs also revealed a superior performance of the ERSPC [28]. For the novel ERSPC-RC the predicted proportion of patients with significant PCa was closer to the observed proportion and the AUC for the detection of significant PCa was significantly higher compared to the PCPT-RC 2.0. Furthermore, the DCAs revealed a somewhat higher clinical net benefit using the ERSPC-RC for significant PCa.

Risk prediction of significant PCa is a very important feature of PCa RCs because there is little evidence that in terms of oncological outcomes, detection and treatment of low-risk PCa is beneficial for the patient [29, 30]. Early detection of low risk PCa can reduce quality of life due to side effects of treatment or patient anxiety [31]. Thus, personalized risk assessment using RC calculating the risk of significant and potentially life-threatining PCa is increasingly proposed to counsel patients prior to prostate biopsies and eventually prevent overdetection and overtreatment [32, 33]. In clinical practice, when given a probability of finding cancer (or significant cancer) by a RC, it should be decided beforehand what threshold this probability will need to reach in order to consider biopsy (i.e. the threshold probability should be set). This decision is often based on patient co-morbidities, and life-expectancy, patient preference and risk tolerance. Our data suggests if the threshold probability for significant PCa at which one would consider biopsy is between 8-35%, then

the ERSPC calculator provides the greatest net benefit, compared to the PCPT risk calculator, the "biopsy all" strategy, and the "biopsy nobody" strategy (see Figure 4). We expect that for most physicians and most patients, a threshold range between 8 and 35% is reasonable to decide whether to perform a biopsy or not, which supports the usefulness of this risk calculator. When predicting risk of cancer overall, it is less clear-cut which risk calculator provides greater net benefit. Nonetheless, in the era of active surveillance, overall cancer risk is a less important trigger for biopsy since Gleason 6 disease would often be managed expectantly. Therefore, the decision of whether or not to biopsy, should ideally be driven by the probability of harbouring significant cancer, where the ERSPC RC was superior to the PCPT 2-0 (AUCs and decision curve analyses).

Our analysis revealed a significant improvement of the novel PCPT-RC 2.0 compared to the PCPT-RC 1.0 with and without prostate volume which constantly overestimated the risk of PCa and significant PCa. However, the PCPT-RC 1.0 with volume showed the highest AUC for significant PCa of the three PCPT-RCs. It has previously been shown that prostate volume is an important variable for the risk prediction of PCa [34]. All ERSPC-RCs include prostate volume either in form of TRUS volume or volume estimates based on DRE [18]. The superior discriminative ability for significant PCa of the ERSPC-RC and the PCPT-RC-1.0 with volume indicates the importance of incorporating prostate volume into PCa risk prediction tools.

There are limitations to our study. First, this a retrospective, single-institution study. Second, we had to perform adjustments that might have an impact on the results of our calculations. Dichotomous outcomes were used for the PCPT-RC 2.0 (rather than three-category outcomes). Also, prostate volume estimates in our database were based on TRUS rather than DRE (ERSPC-RC). Thus, the ERSPC-RC needs further evaluation in studies

using true DRE estimates instead of TRUS-measured prostate volumes. Third, the homogeneous ethnic composition of our cohort might limit the generalizability of our validation.

Conclusion

The use of RCs should be favoured over simple PSA and DRE-based stratification for patient counselling and clinical decision making. Our independent external validation revealed that both updated RC performed better than PSA alone and predicted significant PCa better than any PCa. Prediction of significant PCa, which is important to reduce overdiagnosis and overtreatment was superior using the ERSPC-RC. Despite modifications of both RCs, risk prediction is still not ideal. Implementation of imaging or novel biomarkers might improve the performance of PCa RCs in the future.

Acknowledgments

We would like to thank Monique Roobol for the critical review of the manuscript.

References

[1] Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. Journal of the National Cancer Institute. 2009;101:374-83.

[2] Schroder FH, Hugosson J, Carlsson S, Tammela T, Maattanen L, Auvinen A, et al. Screening for prostate cancer decreases the risk of developing metastatic disease: findings from the European Randomized Study of Screening for Prostate Cancer (ERSPC). European urology. 2012;62:745-52.

[3] Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet. 2014;384:2027-35.

[4] Morote Robles J, Ruibal Morell A, Palou Redorta J, de Torres Mateos JA, Soler Rosello A. Clinical behavior of prostatic specific antigen and prostatic acid phosphatase: a comparative study. European urology. 1988;14:360-6.

[5] Loeb S, Bjurlin MA, Nicholson J, Tammela TL, Penson DF, Carter HB, et al. Overdiagnosis and overtreatment of prostate cancer. European urology. 2014;65:1046-55.

[6] Sakr WA, Grignon DJ, Crissman JD, Heilbrun LK, Cassin BJ, Pontes JJ, et al. High grade prostatic intraepithelial neoplasia (HGPIN) and prostatic adenocarcinoma between the ages of 20-69: an autopsy study of 249 cases. In Vivo. 1994;8:439-43.

[7] Nam RK, Kattan MW, Chin JL, Trachtenberg J, Singal R, Rendon R, et al. Prospective multi-institutional study evaluating the performance of prostate cancer risk calculators. J Clin Oncol. 2011;29:2959-64.

[8] Kranse R, Roobol M, Schroder FH. A graphical device to represent the outcomes of a logistic regression analysis. Prostate. 2008;68:1674-80.

[9] Thompson IM, Ankerst DP, Chi C, Goodman PJ, Tangen CM, Lucia MS, et al. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. Journal of the National Cancer Institute. 2006;98:529-34.

[10] Roobol MJ, Steyerberg EW, Kranse R, Wolters T, van den Bergh RC, Bangma CH, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. European urology. 2010;57:79-85.

[11] Schroder F, Kattan MW. The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review. European urology. 2008;54:274-90.

[12] Lundon DJ, Kelly BD, Foley R, Loeb S, Fitzpatrick JM, Watson RW, et al. Prostate cancer risk assessment tools in an unscreened population. World journal of urology. 2014.

[13] Oliveira M, Marques V, Carvalho AP, Santos A. Head-to-head comparison of two online nomograms for prostate biopsy outcome prediction. BJU Int. 2011;107:1780-3.

[14] Trottier G, Roobol MJ, Lawrentschuk N, Bostrom PJ, Fernandes KA, Finelli A, et al. Comparison of risk calculators from the Prostate Cancer Prevention Trial and the European Randomized Study of Screening for Prostate Cancer in a contemporary Canadian cohort. BJU Int. 2011;108:E237-44.

[15] Cavadas V, Osorio L, Sabell F, Teves F, Branco F, Silva-Ramos M. Prostate cancer prevention trial and European randomized study of screening for prostate cancer risk calculators: a performance comparison in a contemporary screened cohort. European urology. 2010;58:551-8.

[16] Ankerst DP, Boeck A, Freedland SJ, Thompson IM, Cronin AM, Roobol MJ, et al. Evaluating the PCPT risk calculator in ten international biopsy cohorts: results from the Prostate Biopsy Collaborative Group. World journal of urology. 2012;30:181-7.

[17] Ankerst DP, Hoefler J, Bock S, Goodman PJ, Vickers A, Hernandez J, et al. Prostate Cancer Prevention Trial risk calculator 2.0 for the prediction of low- vs high-grade prostate cancer. Urology. 2014;83:1362-7.

[18] Roobol MJ, van Vugt HA, Loeb S, Zhu X, Bul M, Bangma CH, et al. Prediction of prostate cancer risk: the role of prostate volume and digital rectal examination in the ERSPC risk calculators. European urology. 2012;61:577-83.

[19] Foley RW, Lundon DJ, Murphy K, Murphy TB, Galvin DJ, Watson RW. Predicting prostate cancer: analysing the clinical efficacy of prostate cancer risk calculators in a referral population. Irish journal of medical science. 2015.

[20] S. VB, K. G-O. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011;Vol. 45.

[21] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837-45.

[22] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Medical decision making : an international journal of the Society for Medical Decision Making. 2006;26:565-74.

[23] Nam RK, Saskin R, Lee Y, Liu Y, Law C, Klotz LH, et al. Increasing hospital admission rates for urological complications after transrectal ultrasound guided prostate biopsy. J Urol. 2013;189:S12-7; discussion S7-8.

[24] Katz DA, Jarrard DF, McHorney CA, Hillis SL, Wiebe DA, Fryback DG. Health

perceptions in patients who undergo screening and workup for prostate cancer. Urology. 2007;69:215-20.

[25] Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98:691-8.

[26] Auprich M, Augustin H, Budaus L, Kluth L, Mannweiler S, Shariat SF, et al. A comparative performance analysis of total prostate-specific antigen, percentage free prostate-specific antigen, prostate-specific antigen velocity and urinary prostate cancer gene 3 in the first, second and third repeat prostate biopsy. BJU Int. 2012;109:1627-35.

[27] Roobol MJ, van der Cruijsen IW, Schroder FH. No reason for immediate repeat sextant biopsy after negative initial sextant biopsy in men with PSA level of 4.0 ng/mL or greater (ERSPC, Rotterdam). Urology. 2004;63:892-7; discussion 7-9.

[28] Louie KS, Seigneurin A, Cathcart P, Sasieni P. Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO. 2015;26:848-64.

[29] Godtman RA, Holmberg E, Khatami A, Stranne J, Hugosson J. Outcome following active surveillance of men with screen-detected prostate cancer. Results from the Goteborg randomised population-based prostate cancer screening trial. European urology. 2013;63:101-7.

[30] Hayes JH, Ollendorf DA, Pearson SD, Barry MJ, Kantoff PW, Lee PA, et al. Observation versus initial treatment for men with localized, low-risk prostate cancer: a cost-effectiveness analysis. Annals of internal medicine. 2013;158:853-60.

[31] Sanda MG, Dunn RL, Michalski J, Sandler HM, Northouse L, Hembroff L, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. The New England journal of medicine. 2008;358:1250-61.

[32] Roobol MJ. Prostate cancer: Rescreening policies and risk calculators. Nature reviews Urology. 2014;11:429-30.

[33] Thompson IM, Jr., Leach RJ, Ankerst DP. Focusing PSA testing on detection of high-risk prostate cancers by incorporating patient preferences into decision making. Jama. 2014;312:995-6.

[34] Roobol MJ, Schroder FH, Hugosson J, Jones JS, Kattan MW, Klein EA, et al. Importance of prostate volume in the European Randomised Study of Screening for Prostate Cancer (ERSPC) risk calculators: results from the prostate biopsy collaborative group. World journal of urology. 2012;30:149-55.

LEGENDS TO FIGURES

Figure 1 (A-B): Calibration plots for the ERSPC-RC (A) and the PCPT-RC 2.0 (B) predicting any prostate cancer.

The x-axis shows predicted probabilities by the models and y-axis shows observed quintiles. The dashed line represents perfect predictions. The solid line refers to predicted vs. observed event rates with grouped patients (quintiles shown by triangles with horizontal lines as 95% CI). The spikes along the x-axis depict the number of patients with and without prostate cancer.

Figure 2 (A-B): Calibration plots for the ERSPC-RC (A) and the PCPT-RC 2.0 (B) predicting significant prostate cancer. For figure explanation see Figure 1.

Figure 3: Decision curve analysis for the prediction of any PCa on biopsy using the ERSPC-RC (black dashed line) and the PCPT-RC 2.0 (red dashed line). Decision curves examine the theoretical relationship between the threshold probability of PCa biopsy outcome and the relative value of false-positive and false-negative results to determine the value (net benefit) of a predictive model [22]. The horizontal line along the x-axis assumes that no patient will have PCa (ie, no patient should undergo a prostate biopsy), whereas the solid gray line assumes that all patients will have PCa (ie, all patients will need to undergo prostate biopsy).

Figure 4: Decision curve analysis for the prediction of significant PCa on biopsy using the ERSPC-RC (black dashed line) and the PCPT-RC 2.0 (red dashed line). The horizontal line along the x-axis assumes that no patient will have significant PCa, whereas the solid gray line assumes that all patients will have significant PCa.

LEGENDS TO SUPPLEMENTAL FIGURES

Figure 1 (A-B): Calibration plots for the PCPT-RC 1.0 (A) and PCPT-RC 1.0 with prostate volume (B) predicting any prostate cancer.

The x-axis shows predicted probabilities by the models and y-axis shows observed quintiles. The dashed line represents perfect predictions. The solid line refers to predicted vs. observed event rates with grouped patients (quintiles shown by triangles with horizontal lines as 95% CI). The spikes along the x-axis depict the number of patients with and without prostate cancer.

Figure 2 (A-B): Calibration plots for the PCPT-RC 1.0 (A) and PCPT-RC 1.0 with prostate volume (B) predicting significant prostate cancer. For figure explanation see Figure 1.

TABLES

Table 1 Characteristics of the novel ERSPC and PCPT risk calculators

Risk calculator	Description	Variables
ERSPC[1]	any and significant[2] PCa	PSA, prostate volume (categorical), DRE
PCPT-2.0	no PCa, insignificant[3] and significant[2] Pca	PSA, DRE, family history, previous negative biopsy, age, race

[1]RC 3 for biopsy-naïve patients, RC 4 for men with previous negative biopsy

[2]defined as Gleason ≥ 7

[3]defined as Gleason Score 6

Abbreviations: PCa: prostate cancer, PSA: prostate-specific antigen, DRE: digital rectal examination

Table 2: Clinical characteristics of patient cohort and differences between patients with and without a positive prostate biopsy

Variable		All Patients	Negative Biopsy	Positive Biopsy	P-value*
Number of patients		1996 (100)	1513 (76)	483 (24)	
Age at biopsy (years)		63 (58 - 67)	62 (57 - 67)	65 (60 - 69)	**<0.001**
PSA (ng/ml)		5.0 (3.4 -7.6)	4.8 (3.3 - 7.3)	5.6 (3.7 – 9.0)	**<0.001**
Suspicious DRE		386 (19)	241 (16)	145 (30)	**<0.001**
Prostate volume (mL)		40.0 (30.0 - 55.0)	40.0 (30 - 55)	36.0 (28 - 50)	**<0.001**
	<30cm^3	429 (22)	302 (20)	127 (27)	
	30-49cm^3	860 (44)	643 (43)	217 (47)	
	≥50cm^3	654 (34)	535 (36)	119 (26)	
Family history of prostate cancer		40 (2)	21 (1)	19 (4)	**<0.001**
Race	Caucasian	1984 (99)	1504 (99)	480 (99)	>0.99
Prior negative biopsy		845 (42)	707 (47)	138 (29)	**<0.001**
Total cores taken at biopsy	6	139 (7)	117 (8)	22 (4)	**<0.001**
	8	777 (39)	633 (42)	144 (30)	
	12	1080 (54)	763 (50)	317 (66)	
Prostate cancer		483 (24)	-	-	
Significant prostate cancer[1]		226 (11)	-	-	

Values are presented as median (interquartile range) or number (percent)

* P-values in bold indicate a statistically significant difference between the two groups (negative biopsy vs. positive biopsy).

Abbreviations: PSA: prostate-specific antigen, DRE: digital rectal examination

Table 3: Validation results

	ERSPC	**PCPT- 2.0**
Any prostate cancer		
Observed proportion (%)	24	24
Predicted proportion (%)	22.2	24.3
Calibration slope	0.47	-[1]
AUC	0.65	0.66
Significant prostate cancer		
Observed proportion (%)	11	11
Predicted proportion (%)	9.4	8.6
Calibration slope	0.56	-[1]
AUC	0.73	0.70

AUC: Area under the receiver operating characteristic curve
[1] The PCPT-2.0 risk calculator is based on a multinomial logistic regression model instead of a binary logistic regression model